



Superfast Superhighway

We may be used to stuttering internet connections but, says **Jon Thompson**, these may soon be a thing of the past. **FAST** promises to boost internet speeds dramatically, and we won't even need to pay a penny

Imagine being able to download enough data from the internet in 30 seconds to fill a DVD. This may soon be possible according to researchers in California, who claim that the way the internet currently works wastes up to 75 per cent of its capacity. What's more, they have worked out how to reclaim that lost capacity, and most users won't even have to change a thing.

The internet has coped well since its popularity exploded. In the past 15 years it has successfully scaled up by an incredible six orders of magnitude, and that growth shows no sign of abating. At the end of June 2005, the web-watchers at Netcraft (www.netcraft.com) announced that in that month alone, 2.76 million new host names had appeared. By design, however, the same mechanism that sustains this continued expansion also makes the internet much slower than it could be.

Originally designed by Vint Cerf and Bob Kahn in the mid-1970s, the internet's underlying Transmission Control Protocol (TCP) ensures that packets of data arrive safely. To do this, each packet has a unique, sequential number. When one computer sends a packet to another, the receiver sends back an acknowledgement quoting this number to say that it arrived safely. If the sender doesn't hear back after a certain period of time, it assumes there's a problem and sends it again.

In terms of speed this mechanism is TCP's weakness. If an acknowledgement fails to arrive after a set period of time, TCP doubles the time it waits before sending further packets to help clear

whatever congestion has caused the problem. If it still fails to receive a timely acknowledgement, it doubles the interval again, and so on. When acknowledgements start to arrive quickly again, TCP increases the transmission rate only gradually. This over-cautious approach is frustratingly apparent to modem and broadband users, because even a slight delay causes vast drops in the data transfer rate. This is why you might start a lengthy download with a healthy throughput only to see it fall to a mere trickle within minutes.

On ultra-high-speed links such as those that serve the scientific community, the problem is far worse. If you need to send massive amounts of data, transmission delays caused by slow acknowledgement packets can quickly stretch data

transfer times to several days. The current solution is to add more or faster links, but wouldn't it be better to rethink how we use TCP and unlock that hidden 75 per cent?

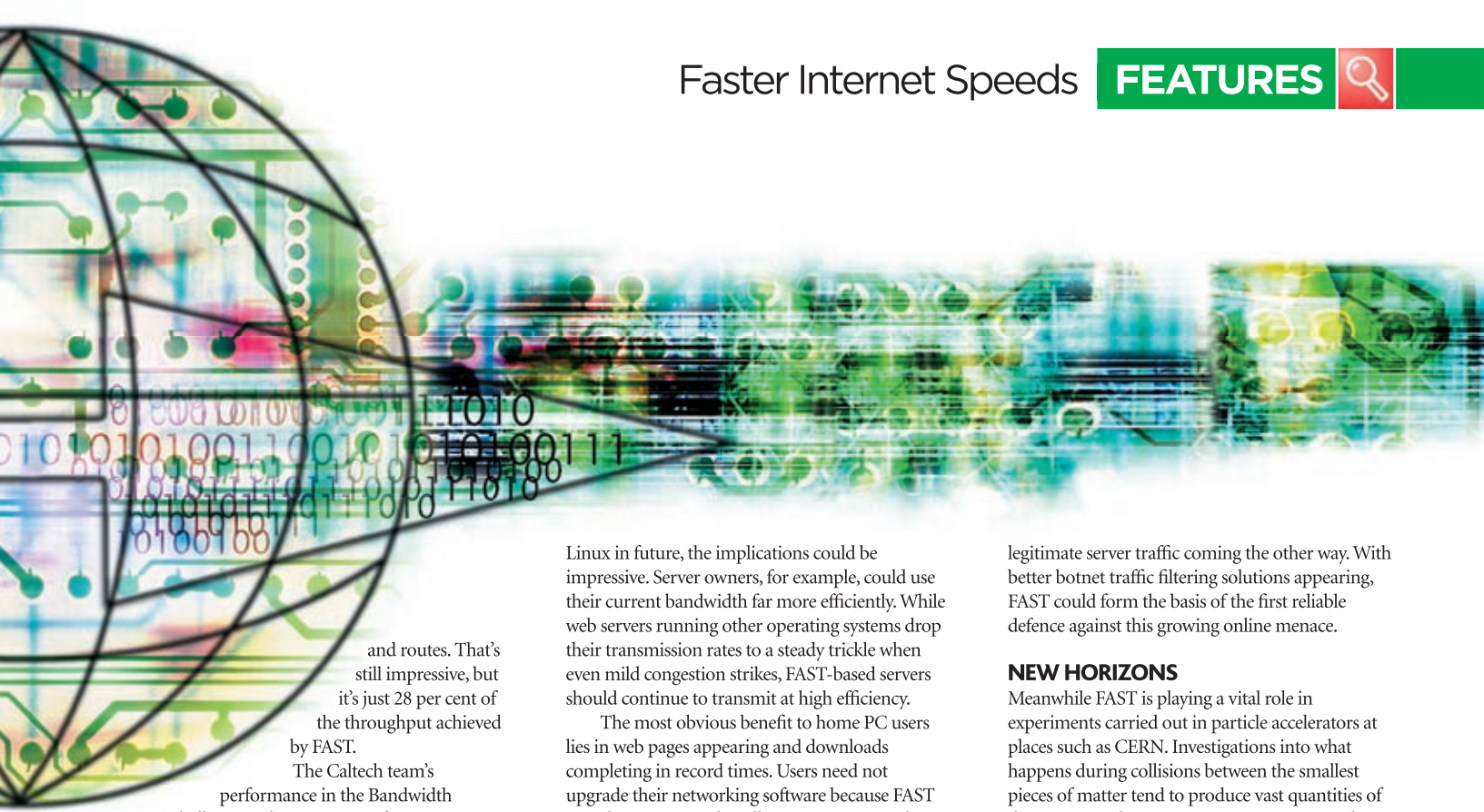
TOP SPEED

Associate Professor Steven Low and his team at Caltech, the California Institute of Technology, thought so. In 2003, they started to develop a novel way of using TCP called fast active queue management scalable transmission control protocol (FAST) – see their website at <http://netlab.caltech.edu/fast>. FAST is compatible with TCP but makes intelligent decisions about congestion levels to maximise the number of packets it transmits. The team has achieved remarkable levels of efficiency in the lab. "Over a given internet link," says Low, "the current protocol tops out at 25 per cent maximum throughput. FAST is 95 per cent efficient."

In proving this claim, Low's team first smashed internet speed records set by others and has now started breaking its own. During 2003's SC2003 Supercomputing Conference Bandwidth Challenge competition in Baltimore, the Caltech team, aided by researchers from Stanford University and the European particle physics lab CERN (www.cern.ch) successfully transmitted data 10,000km across the internet from California to the CERN data centre near Geneva. They managed a peak throughput of 925Mbit/s. Conventional TCP managed just 266Mbit/s using the same connections, equipment



At 2004's Supercomputer Conference in Pittsburgh, an international consortium headed by Caltech's FAST team set a world internet speed record of just over 101Gbit/s



and routes. That's still impressive, but it's just 28 per cent of the throughput achieved by FAST.

The Caltech team's performance in the Bandwidth Challenge at the SC2004 conference was even more impressive and won them the Sustained Bandwidth award. Here the FAST team (plus a large, worldwide group of collaborators) presented a demonstration called High-Speed Terabyte Transfers for Physics, during which they succeeded in transmitting a whopping 101Gbit/s across the conference floor to the Fermilab booth and to an external site in Los Angeles. That's the equivalent of three full DVDs of data every second, or the entire US Library of Congress every 15 minutes.

FAST's performance during SC2004 was greater than that of all the entries in the previous year's competition put together. To achieve this level of performance, the FAST team needed links, routers and even network cards that could handle the amount of data they wanted to send, including multiple 10Gbit/s links working in parallel and connected to Cisco 7600 and 6500 carrier class routers. During the demonstration, National Lambda Rail, suppliers of some of the external links, found that FAST used up to 99 per cent of the available bandwidth.

The team has an open-source beta version of its software for Linux on its website. Access is currently restricted to non-commercial development partners and educational establishments, but once it's stable the team plans to make it available free for public use. Once this happens, we may see the advantages very quickly.

SWITCHING TO FAST

Although nothing can make hardware transmit or receive data faster than its physical limitations, the Caltech team's work has the potential to accelerate both domestic broadband and dial-up connections. FAST causes computers to transmit more data per second using the available bandwidth, so devices on the receiving end don't have to do anything special. There's no possibility of overloading the receiver as, like conventional TCP, FAST adjusts its performance according to how quickly it receives acknowledgements. For an explanation, see the box on page 195.

Well over half of the ever-increasing number of public web servers run Apache, most of them under Linux, on which FAST can already run. If FAST or something like it comes bundled with

Linux in future, the implications could be impressive. Server owners, for example, could use their current bandwidth far more efficiently. While web servers running other operating systems drop their transmission rates to a steady trickle when even mild congestion strikes, FAST-based servers should continue to transmit at high efficiency.

The most obvious benefit to home PC users lies in web pages appearing and downloads completing in record times. Users need not upgrade their networking software because FAST uses the same TCP that all internet-connected computers use, but in a more intelligent way. This is because, aside from the odd task such as sending an email with an attachment, little traffic flows from client to server. When a user clicks on a link, he might send a couple of kilobits containing the requested URL, but he might receive back several hundred kilobytes of traffic as the HTML, Java, images and so on flow to his browser.

If FAST-like techniques are built into domestic operating systems, peer-to-peer and online game enthusiasts will see speed benefits, too, enabling them to transmit at high speed as well as receive. Still, there are advantages to leaving users' PCs as they are. By not incorporating FAST into home operating systems, so-called botnets (networks of poorly patched systems that are infected and then used to attack websites or send out spam) will suffer. If home computers stick to the old-style TCP, botnets should run a lot slower than

legitimate server traffic coming the other way. With better botnet traffic filtering solutions appearing, FAST could form the basis of the first reliable defence against this growing online menace.

NEW HORIZONS

Meanwhile FAST is playing a vital role in experiments carried out in particle accelerators at places such as CERN. Investigations into what happens during collisions between the smallest pieces of matter tend to produce vast quantities of data. Once under way, some experiments produce terabytes (1,024GB) and even petabytes (1,024TB) of data almost instantly. Collaborators on such experiments, who are spread around the world, need to analyse the results.

According to the FAST team, experiments by the High Energy and Nuclear Physics (HENP) community can involve over 2,000 physicists at 150 universities in 30 countries. The amount of data involved is a serious impediment to collaboration.

These days, storage is no problem, but moving this amount of data around is extremely difficult. At 200Kbit/s it would take 1,395 years to transmit a petabyte of data. Burning it to CD wouldn't help either, as you'd need 1.53 million discs. "We need a global network capable of sending these files around the world, not just across a university," says FAST team leader Professor Low.

The problem is about to get much worse for the physicists at CERN. In 2007, experiments



Previous attempts at going faster

Bandwidth is expensive, which is why there have been many attempts to use it more efficiently over the years. Rather than making intelligent decisions about the transmission rate, as FAST does, these have focused on routing traffic or prioritising its flow by type.

The internet is fault-tolerant and provides many potential paths from source to destination. Each step in those paths is controlled by routers. Rather than working in blind isolation, routers swap congestion information in a number of ways, such as using the router information protocol. This ensures that, as far as the routers involved are aware, all links are operating as efficiently as possible. This is OK at a raw traffic level, but it doesn't address the urgency of individual packets.

Buried in the header of every transmitted packet are bits that describe the Quality of Service (QoS) it should receive. The QoS bits tell routers which packets to send first if they have a queue. Over the years, a number of products have used

QoS to favour the transmission of some packets over those deemed less important. Streaming video has a higher priority than email, for instance.

Linux and Windows XP now carry the ability to set QoS bits, but this is open to abuse. One user might decide that their application is more important than everyone else's and bump up the priority on its packets. To overcome this, some current prioritisation systems wait until the packets are out on the local network before stamping them with the proper QoS settings.

An example of this is Packetshaper by network management specialists Packeteer (www.packeteer.com). It sits on a network just behind its internet connection. Designed for bandwidth-hungry companies, it sets the QoS bits on packets as they flow by, based on source, destination IP address and port number. One advantage of accessing all packets is that Packetshaper can collect and display information about the nature and make-up of the traffic.



IPv6

The current internet protocol, IPv4, provides 4.3 billion unique addresses. Twenty-five years ago, its designers thought this more than anyone would ever need. Now there are only around a billion left, and the rate of demand is increasing. "A billion sounds like a hell of a lot," says Martin McNealis, director of router giant Cisco. "But places like China and India are now coming online. People are talking about one billion home telephones getting IP addresses. It starts to sound smaller and smaller."

The US holds the rights to 70 per cent (three billion) of the total IPv4 addresses. This is to do with assumptions made 25 years ago about the growth of new internet connections and where they would occur. Other parts of the world have not been quite so lucky. India, while experiencing massive growth in internet access, has an allocation of just two million addresses serving an eager population of over a billion.

Japan, where internet-enabled devices are part of everyday life, has already started abandoning IPv4 out of necessity, and broadband-saturated South Korea is not far behind. With FAST TCP promising to free up vast amounts of hidden bandwidth for more internet-enabled devices, demand for addresses could skyrocket as new applications become possible.

The solution is IPv6, which makes more addresses available than there are grains of sand in the Sahara: 340 billion billion billion. It also uses strong encryption to secure individual packets. This was key to the US Department of Defense's endorsement of the standard in October 2003. It currently runs the world's largest multi-vendor test of an IPv6 network, dubbed Moonv6 (<http://moonv6.sr.unh.edu>), and plans to transfer its Global Information Grid Network to the new protocol by 2008.

In Asia many hope that security-minded US companies will see the sense in adopting IPv6. Cody Christman, product engineering director at Japanese phone giant NTT DoCoMo, said: "With the Defense Department's announcement, a lot of CTOs and CIOs realise that IPv6 is going to come [to the US]. The push is going to come." Verio, a subsidiary of NTT, is taking this message to corporate America. To convince US IT directors of the benefits of IPv6 networking, Christman says that Verio is to make IPv6 available in San Francisco, Los Angeles and Washington, DC.

In the end, though, it could be home users who push hardest for the switch to IPv6. In an increasingly wired but security-conscious world, transmitting data over IPv4 may soon be seen as unacceptably insecure. Microsoft Windows XP is already capable of running IPv6, as are new Linux distributions and Apple's Mac OS X. In the next couple of years we may start to see ISPs gain market share by offering IPv6 connections to address security fears. When IPv6 is finally adopted worldwide, in conjunction with FAST-like TCP techniques, it will make possible the biggest increase in capacity that any communications medium has seen since the birth of radio.

running on the new Large Hadron Collider as part of the Compact Muon Solenoid Collaboration project begin. This time, the scientists are looking for the highly illusive Higgs particles that are thought to be the cause of mass in the universe. The data produced will run into the Exobyte range (a one followed by 18 zeros), and project engineers say they will need a sustained throughput of at least 10Gbit/s, 24 hours a day, seven days a week to distribute it to collaborators.

Others are about to benefit from FAST, too. The improved Internet2 (www.internet2.edu), a high-speed, subscription-only network currently under development for use between research universities and collaborators in industry, uses standard TCP to run at about 350Mbit/s over dedicated fibre-optics.

The hope is that FAST will triple this to 1Gbit/s using the same infrastructure. At this speed, a petabyte of data takes 97 days to transmit. This is still a long time, but it's considerably faster than conventional TCP. At Caltech's record speed of 101Gbit/s, this would fall to just over 23 hours.

Researchers at Stanford University's Linear Accelerator Center in California (www.slac.stanford.edu) have much the same problem and need such a network when transmitting the results of their atom-smashing activities. They used FAST in speed trials and created a private test network from dedicated fibre-optic links and \$1million (around £568,000) of equipment loaned by router giant Cisco.

In one experiment, they sent the equivalent of two DVD movies 6,800 miles from California to Amsterdam in just 58 seconds. "Imagine being able to download two full-length, two-hour movies within a minute. That changes the whole idea of how media is distributed," enthuses Les Cottrell, assistant director of the Stanford accelerator.

THAT'S ENTERTAINMENT

Microsoft and Disney are enthusiastic about streamed entertainment and already have a very strong interest in the commercial potential of FAST. Each is understood to be in talks with Caltech about how it might use the technology to provide video-on-demand services over the internet. Microsoft has already started moving into the home entertainment market with its Windows XP Media Center Edition.

Late in the summer of 2003, Disney announced its MovieBeam video-on-demand system. The stage seemed set for a VHS against Betamax-style showdown, and many industry watchers said they expected Microsoft to emerge not only as the winner, but also as a serious global digital media company.

As if fulfilling these predictions, Disney withdrew its MovieBeam service, which had been operating in three US cities, earlier this year. The associated website says that the company is "updating and expanding the best on-demand movie service in America". Critics say that it has closed for good, because the quality of the video being streamed to its set-top boxes was below that of DVD, and the service was too expensive at \$7 (around £4) per month and \$2.50-\$4 (£1.42-£2.27) per download. If it is to provide full DVD-quality movies the provider can't skimp on bandwidth, and the increasing ubiquity of the internet makes it ideal as a high-quality delivery mechanism. The only problem until now has been its speed.

SWAP SHOP

Meanwhile, the Motion Picture Association of America (MPAA) has got wind of these advances. Its great fear, fuelled by rumours of file swapping, is that FAST and Internet2 users could trade pirated material very quickly, despite the fact that Internet2 is a subscription-only service for scientific collaboration. As a result, the MPAA has controversially requested membership of the Internet2 scheme to monitor for any such activities and prosecute offenders accordingly.

The MPAA may have a point. In 2002, Joe St Sauver of the University of Oregon published a paper claiming that up to 30 per cent of the throughput of some Internet2 links contained illegal file-swapping traffic. Although further research shows that this figure is now around seven per cent, the problem persists and is far more prevalent during term time.

Chris Russell, the MPAA's vice-president of internet standards and technology, says his group is also in talks with the Internet2 consortium to explore new ways of delivering legitimate, high-quality content to the domestic consumer. "We've been working with Internet2 for a while to explore ways we can take advantage of delivering content at these extremely high speeds, and basically manage illegitimate content distribution at the same time."



Experiments on sub-atomic particles such as those conducted in the CERN particle accelerator produce mind-boggling quantities of data, which if shared over a normal TCP connection could take hundreds of years to reach its destination

Photo: CERN



TCP v FAST

Transmission Control Protocol and Internet Protocol ensure that packets of data arrive intact at the right computer. IP is responsible for addressing packets of data, while TCP makes sure all the packets get through.

Consider two users, Bob and Alice. Alice wants to send Bob a file. First the TCP software on Alice's PC sends a request for a connection, called a SYN packet, to the relevant port. If Bob's computer is happy to accept Alice's connection, it sends a SYN/ACK packet back in acknowledgement, which contains a random sequence number. If Alice's computer is happy too, it sends back a final acknowledgement called an ACK. The PCs now know about each other and are happy to communicate on the specified port.

Alice's PC then starts to send packets of information. It numbers each sequentially, starting from the number Bob's computer sent it. To control the flow, Alice's computer uses a TCP parameter called window size to determine how many packets to send before it waits for acknowledgements to come back to confirm that the packets arrived safely.

If acknowledgements don't come back soon enough, Alice's PC assumes the network is congested. To compensate, her PC doubles the delay between transmitting each packet, slowing its transmission rate. If another acknowledgement fails to arrive, it doubles the delay again. When acknowledgements start to arrive at an acceptable rate, Alice's computer gradually increases the transmission rate. This cautious technique is called 'additive-increase, multiplicative-decrease'.

Caltech's FAST TCP researchers decided to see what happens when congestion strikes a TCP network. They found that as various computers react overcautiously to bottlenecks, network throughput falls dramatically and ends up far below its potential maximum. This gave them an idea of how to use TCP more efficiently.

Rather than blindly doubling the delays when acknowledgements arrive slowly, FAST TCP starts by measuring the delay between sending a packet and receiving the acknowledgement to see whether or not it is late. From this it can work out the maximum number of packets it can transmit safely, which is far more than traditional TCP assumes. This cautious approach to slowing the transmission rather than to transmitting data is the key to FAST's impressive performance.

FAST consists of four components. The data control module decides which packets to send next. When packets are lost, data control decides when to retransmit them. Next, the window control module determines the size of the TCP window (the number of packets it transmits at once) based on what it knows about the round-trip time between Alice and Bob.

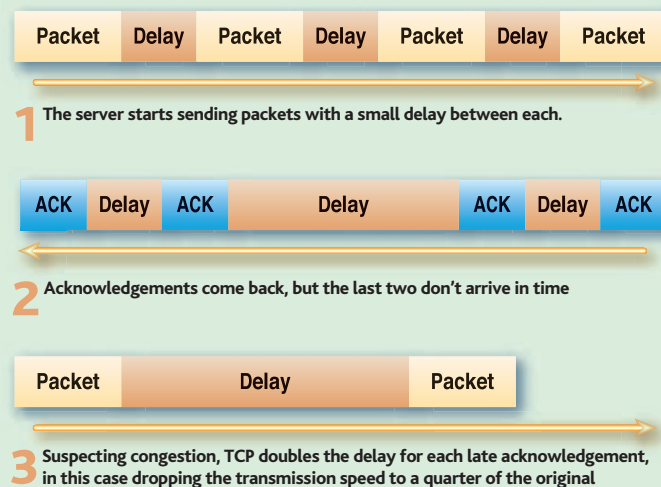
The 'Burstiness' control module decides on the best time to send these packets, avoiding sudden bursts of activity that may cause congestion. It does this by tracking the average delay times packets experience, which enables it to make better use of the bandwidth than TCP.

The estimation module supplies congestion information. This uses rather complex equations to work out the average congestion delay, the actual delays experienced by packets and when to declare overdue packets lost. The other modules use this to determine which packets to re-send and when to send them.

Together, these modules make intelligent decisions based not only on what's happening on the network, but also on how it has reacted in the past. In this way, FAST can get an idea of how much data it can safely send and which congestion information it can safely ignore.

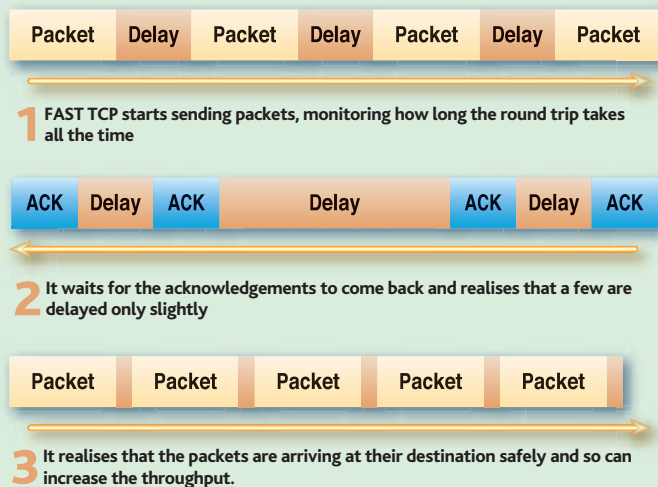
TCP

TCP is designed to ensure data gets through, but it doubles the delay between packets if acknowledgements are slow in coming back



FAST

By monitoring the link constantly rather than simply abandoning packets with late acknowledgements, FAST TCP can safely increase throughput



High-speed piracy doesn't just bother the MPAA. Other groups are increasingly worried, too. "The speed of these networks... allows users to obtain copyrighted movies in minutes and music in seconds," claimed Cary Sherman, president of the Recording Industry Association of America, in evidence to Congress in October 2004. "Further, the closed nature of these networks, being available only to those engaged in academia, makes it more difficult for copyright owners to protect their works and to notify responsible parties of their infringement."

The public membership list for the Internet2 consortium reveals some surprising names. Along with Microsoft and Warner Brothers are the likes of Napster, so it may not be long before we start seeing the first services based on high-speed downloads.

FULL SPEED AHEAD?

At the moment, TCP trades speed for reliability. Using FAST-like techniques, this bottleneck will shift to the receiving computer, which is something that bothers Caltech physics professor Harvey Newman. "You have this inversion where the limitations on advances will not be the speed of the internet but rather the speed of your computer," he says.

New consumer devices, designed with faster processors and with higher data throughput in mind, will not suffer in this way, however. The revolutionary new Cell microprocessor, for instance, which was designed jointly by IBM, Sony and Toshiba, can handle high data rates. It's also cheap enough for use in games consoles and internet-enabled set-top boxes.

As an example of Cell's promise Toshiba demonstrated a system at the 2005 Cool Chips show in Japan that used a single Cell chip to decode and display 48 video streams simultaneously. Apart from powering the upcoming PlayStation 3, plans for Cell include a range of handheld communication devices, as well as Linux servers and home entertainment products.

It looks as if we're in for a high-speed future, but what else will we find to do with it? Voice over IP internet telephony is taking off, as is the next-generation internet protocol IPv6 (see the box opposite), which will provide enough addressing capacity to internet-enable millions of new smart devices. If FAST can raise the internet's speed limit, all these components will fall into place to make this a truly connected world. ☑